AFIT/GCS/ENG/99M-16

# INCORPORATING SCENE MOSAICS AS VISUAL INDEXES INTO UAV VIDEO IMAGERY DATABASES

THESIS

Timothy I. Page,
Captain, USAF

AFIT/GCS/ENG/99M-16

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



**1. AGENCY USE ONLY** *(Leave blank)*

**2. REPORT DATE**
March 1999

**3. REPORT TYPE AND DATES COVERED**
Master's Thesis

**4. TITLE AND SUBTITLE**
Incorporating Scene Mosaics as Visual Indexes into UAV Video Imagery Databases

**5. FUNDING NUMBERS**

**6. AUTHOR(S)**
TIMOTHY I. PAGE, Capt, USAF

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
Air Force Institute of Technology
WPAFB, OH 45433-7765

**8. PERFORMING ORGANIZATION REPORT NUMBER**
AFIT/GCS/ENG/99M-16

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
Maj Steven M. Matechik
AFRL/IFEC
26 Electronic Parkway
Rome, NY 13441-4514

**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**

**11. SUPPLEMENTARY NOTES**
Maj Michael L. Talbert (advisor)
michael.talbert@afit.af.mil
DSN 785-6565 ext 4280 COMM (937) 255-6565

**12a. DISTRIBUTION AVAILABILITY STATEMENT**
Approved for public release; distribution unlimited

**12b. DISTRIBUTION CODE**

**13. ABSTRACT** *(Maximum 200 words)*
The rise of large digital video archives has strengthened the need for more efficient ways of indexing video files and accessing the information contained in them. Reconnaissance platforms, such as the Predator UAV, are contributing thousands of hours of video footage that require analysis, storage, and retrieval. A process is proposed for converting a video stream into a series of mosaic and selected still images that provide complete coverage of the original video. The video mosaic images can be utilized as visual indexes into a video database. In addition, mosaic images contain information from an entire sequence of video frames to provide "at a glance" analysis capabilities. Actual reconnaissance video footage is converted to still-image representation using the proposed process and the results are discussed. Further, a web-based browse and search capability was developed to demonstrate the benefits of using the proposed process. Further, the Predator Unmanned Aerial Vehicle (UAV) system configuration is described with recommendations for placement of the video mosaic building process proposed in this research.

**14. SUBJECT TERMS**
UAV Full Motion Video, Video Mosaic, Video Segmentation, Video Information Storage and Retrieval

**15. NUMBER OF PAGES**
60

**16. PRICE CODE**

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| Unclassified | Unclassified | Unclassified | UL |

AFIT/GCS/ENG/99M-16

# INCORPORATING SCENE MOSAICS AS VISUAL INDEXES INTO UAV VIDEO IMAGERY DATABASES

THESIS

Presented to the Faculty of the Graduate School of Engineering

of the Air Force Institute of Technology

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science in Computer Systems

Timothy I. Page,

Captain, USAF

March 1999

AFIT/GCS/ENG/99M-16

INCORPORATING SCENE MOSAICS AS VISUAL INDEXES INTO

UAV VIDEO IMAGERY DATABASES

THESIS

Timothy I. Page
Captain, USAF

Approved:

Michael L. Talbert, Major, USAF
Chairman

15 Mar 1999
Date

Dr. Henry B. Potoczny, Ph.D.
Member

15 Mar 99.
Date

Richard Raines, Major, USAF
Member

15 Mar 99
Date

Stephen M. Matechik, Major, USAF
Member

25 Feb 99
Date

# ABSTRACT

The rise of large digital video archives has strengthened the need for more efficient ways of indexing video files and accessing the information contained in them. Reconnaissance platforms, such as the Predator UAV, are contributing thousands of hours of video footage that require analysis, storage, and retrieval. A process is proposed for converting a video stream into a series of mosaic and selected still images that provide complete coverage of the original video. The video mosaic images can be utilized as visual indexes into a video database. In addition, mosaic images contain information from an entire sequence of video frames to provide "at a glance" analysis capabilities. Actual reconnaissance video footage is converted to still-image representation using the proposed process and the results are discussed. Further, a web-based browse and search capability was developed to demonstrate the benefits of using the proposed process. Further, the Predator Unmanned Aerial Vehicle (UAV) system configuration is described with recommendations for placement of the video mosaic building process proposed in this research.

# ACKNOWLEDGMENTS

Many people have contributed their time and expertise to ensure the successful completion of this research effort. I would like to this opportunity thank those who have helped in one way or another.

First, I would like to thank my thesis advisor, Maj Michael Talbert, for your undying patience and wisdom. Your knowledge of databases and enthusiasm toward new technology helped me to "Think out of the box" when developing my thesis topic. Your help and guidance kept me focused on my research with appropriate course corrections now and then.

Next, I would like to thank my sponsor, Maj Stephen Matechik from Rome Laboratory for providing me with the mosaic building software and sponsorship. The information you provided was invaluable to my research.

I would also like to thank the instructors in the Computer Science Department at AFIT with whom I had the pleasure of their instruction. You shared your knowledge and expertise sprinkled with just the right level of humor to keep it enjoyable.

For their support, comradery and friendship, I would like to thank my classmates and peers in GCS-99M. Experiencing AFIT with you next to me has made it a fun experience.

For their love, support and understanding, I would like to thank my family. To Carol, words cannot describe what you mean to me. Your friendship and encouragement helped give me the energy I needed to finish my research. Your sacrifices did not go unnoticed; the nights you took care of Kevin and Chris so that I could get much needed sleep, providing me the time I needed to finish an assignment, Your understanding of the long hours of research. To Kevin and Christyn, thanks for understanding when Daddy had to work. Your hugs and kisses helped me to forget the long hours.

Finally, I would like to thank God, for without God nothing truly worth while is possible.

Timothy I. Page

TABLE OF CONTENTS


1 INTRODUCTION ..... 1

1.1 OVERVIEW ..... 1

1.2 BACKGROUND ..... 3

1.3 PROBLEM ..... 4

1.4 RESEARCH OBJECTIVE ..... 5

1.5 SCOPE ..... 6

1.6 APPROACH AND PRESENTATION ..... 6

2 SURVEY AND ANALYSIS OF PREVIOUS WORK ..... 8

2.1 INTRODUCTION ..... 8

2.2 PREDATOR SYSTEM ..... 8

2.2.1 *Overview* ..... *9*

2.2.2 *Predator Aerial Vehicle* ..... *9*

2.2.3 *Summary of Predator System* ..... *10*

2.3 IMAGE AND VIDEO FORMATS ..... 10

2.3.1 *Joint Photographic Experts Group (JPEG)* ..... *11*

2.3.2 *Portable Pixel Map (PPM)* ..... *11*

2.3.3 *Thumbnail* ..... *12*

2.3.4 *Moving Picture Experts Group (MPEG)* ..... *12*

2.4 VIDEO STREAM SEGMENTATION ..... 14

2.5 BUILDING A VIDEO MOSAIC IMAGE. ..... 15

2.5.1 *Frame Stitcher* ..... *15*

2.5.2 *VideoBrush Panorama* ..... *17*

2.5.3 *Visual Stitcher* ..... *17*

2.5.4 *Video Characteristics that Hamper Mosaic Building* ..... *18*

2.5.4.1 *Zoom* .......... 18

2.5.4.2 *Tilting* .......... 18

2.5.4.3 *Scene Changes/Bad Frames* .......... 19

2.5.4.4 *Summary* .......... 19

2.6 VIDEO/IMAGE STORAGE AND RETRIEVAL .......... 20

2.6.1 ***Video Database Management System (VDBMS)*** .......... 20

2.6.2 ***Visual Information Retrieval (Browsing)*** .......... 21

2.6.3 ***Object Oriented Database*** .......... 22

2.6.4 ***World Wide Web (WWW) Technology*** .......... 22

2.6.4.1 *Web Server* .......... 22

2.6.4.2 *Web Client* .......... 22

2.6.4.3 *HyperText Markup Language (HTML) Page* .......... 22

2.6.4.4 *Active Server Pages (ASP)* .......... 23

2.7 SUMMARY .......... 23

**3 VIDEO MOSAIC BUILDING PROCESS .......... 24**

3.1 INTRODUCTION .......... 24

3.2 PROCESS DESCRIPTION .......... 25

3.2.1 ***Data Collection*** .......... 25

3.2.2 ***Identify Video Sequences to Build Mosaics*** .......... 26

3.2.3 ***Identify Frame to Represent Non-Candidate Video Sequences*** .......... 27

3.2.4 ***Decode MPEG-2 Video Stream*** .......... 27

3.2.5 ***Build Mosaics*** .......... 28

3.2.6 ***Extract Thumbnails for Mosaics and Selected Still Images*** .......... 29

3.2.7 ***Collect Meta Data for Video Stream and By-Products*** .......... 29

3.2.8 ***Place Archival Files on Data Server*** .......... 30

3.2.9 ***Create Static Web Pages*** .......... 31

3.2.10 ***Deploy Web Search Page*** .......... 31

3.2.11 *Serve Web Pages on Demand* ..... 32

3.3 SUMMARY ..... 32

4 ***PROCESS IMPLEMENTATION AND DEMONSTRATION*** ..... **33**

4.1 OVERVIEW ..... 33

4.2 PROCESS IMPLEMENTATION AND DEMONSTRATION DEVELOPMENT ..... 33

4.2.1 *Data Collection* ..... 33

4.2.2 *Identify Video Sequences to Build Mosaics* ..... 34

4.2.3 *Identify Frames to Represent Non-Candidate Video Sequences* ..... 34

4.2.4 *Decode MPEG-2 Video Stream* ..... 34

4.2.5 *Build Mosaics* ..... 35

4.2.6 *Extract Thumbnails for Mosaics and Selected Still Images* ..... 36

4.2.7 *Collect Meta Data for Video Stream and By-Products* ..... 36

4.2.8 *Place Archival Files on Data Server* ..... 37

4.2.9 *Create Static Web Pages* ..... 37

4.2.10 *Deploy Web Search Page* ..... 38

4.2.11 *Serve Web Pages on Demand* ..... 39

4.2.12 *Summary* ..... 39

4.3 VIDEO MOSAIC PROCESS LOCATION IN THE PREDATOR SYSTEM ..... 40

4.3.1 *Ground Control Station* ..... 40

4.3.2 *Archival Location* ..... 40

4.3.3 *Process Location Conclusion* ..... 41

5 ***CONCLUSIONS AND RECOMMENDATIONS*** ..... **42**

5.1 INTRODUCTION ..... 42

5.2 CONCLUSIONS ..... 42

5.3 RECOMMENDATIONS ..... 44

5.4 SUMMARY ........................................................................................................ 44

## LIST OF FIGURES


FIGURE 1-1. VIDEO MOSAIC IMAGE AND SINGLE EXTRACTED STILL FRAME. 3

FIGURE 2-1. PREDATOR SYSTEM OVERVIEW [8] 9

FIGURE 2-2. PREDATOR UNMANNED AERIAL VEHICLE IN FLIGHT. 10

FIGURE 2-3. A SEQUENCE OF MPEG FRAMES[14]. 13

FIGURE 2-4. OVERLAPPING IMAGE ALIGNMENT USING VARYING RESOLUTIONS 16

FIGURE 2-5. EXAMPLE FRAME STITCHER MOSAIC INFORMATION FILE (MIF). 17

FIGURE 2-6. A DEPICTION OF SCENE TILTING EFFECT. 19

FIGURE 2-7. THREE LEVELS OF VIDEO DEGRADATION. 19

FIGURE 2-8. A HIERARCHY OF DIGITAL VIDEO. 21

FIGURE 3-1. OVERALL CONVERSION PROCESS (VIDEO/MOSAIC/THUMBNAIL). 25

FIGURE 3-2. SEGMENTATION OF MPEG-2 VIDEO. 26

FIGURE 3-3. MPEG SEGMENTS DECODED INTO STILL IMAGE FILES. 27

FIGURE 3-4. CREATION OF VIDEO MOSAIC A USING SEQUENCE OF STILL FRAMES. 28

FIGURE 3-5. JPEG IMAGE CONVERTED TO THUMBNAIL IMAGE 29

FIGURE 3-6. STATIC WEB PAGE DISPLAYING MOSAIC IMAGES 31

FIGURE 4-1. STATIC WEB PAGE FOR BROWSING MOSAICS. 37

FIGURE 4-2. QUERY CRITERIA COLLECTION FORM. 38

FIGURE 4-3. SEARCH RESULTS PAGE. 39

# 1 INTRODUCTION

## 1.1 Overview

Today more than ever multimedia information has become a vital part of business, and as well as military operations. The multimedia data being collected is growing at a staggering rate, making it exceedingly difficult to retrieve needed portions. To compound this problem, multimedia data is difficult to handle due to the magnitude of its size.

Technological advances in the areas of computers and communications provide capability for data users to have access to their data wherever they may be. In today's business environment, and particularly in military operations, having timely access to information of all forms is critical. The information needed to make a sales order in a retail business needs to be available, or such a business may not survive in today's competitive business environment. It is even more important that military organizations have access to the information needed to plan a mission or assess its effectiveness. Much of the information needed to plan missions or perform battle damage assessment (BDA) is multimedia data, which has high communications and computer resource requirements. A capability is needed to ensure that information is available in a form that can be efficiently accessed. A new technology that builds a single mosaic image from a video stream can be used to exploit reconnaissance imagery in support of the war fighter. A system that applies this capability could minimize processing times for military intelligence data collection, analysis, and dissemination, possibly making the difference between winning or losing a conflict.

Military intelligence multimedia information often combines video, audio, still images, and text. While today’s database management systems (DBMS) are addressing the

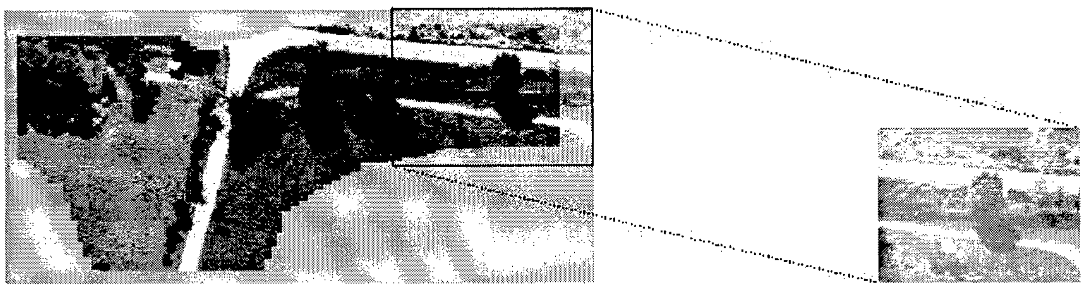
issue of multimedia files, the combination of video, audio, stills, and text results in complex data types. Making the newly emerging multimedia technology easily accessible to the user is a complex task for database developers due to the varying size and formats of the files. Video database servers allow users to find and retrieve video files, however, some military needs go beyond these capabilities. Target selection during mission planning or BDA of a previous target may require access to information contained somewhere inside a large video file. Such information may be contained in only a few frames of the file. Therefore, it is important to index on such a sequence of frames for quick identification and retrieval later.

Parsing streamed video into semantic segments is another key enabling technology. The streaming nature of a video file can make locating a specific frame very difficult and time consuming. Methods have been developed to automatically detect scene changes in a video segment [1, 20, 17]. Incorporating these methods will provide a way to automatically break a video stream into scene-specific segments.

Even with segment detection, the "soda straw" presentation of streamed video reduces its applicability for intelligence analysis. A video segment can be converted into a sequence of still images by capturing individual frames. Adjacent frames will contain redundant data. Processing the consecutive frames through a mosaic building application will reduce the segment to a single composite forming a panoramic image. The size of a mosaic image is inversely proportional to the degree of overlap between successive frames.

Current procedures in the intelligence community use "frame-grabbed" still images that are exploited, reported, and disseminated. The exploited imagery is sent directly to the theater combat operations center and/or to an imagery server for archival [8]. Video mosaic

images provide the advantage of seeing the entire scene as opposed to only a single frame. Figure 1-1A contains a mosaic while Figure 1-1B contains a single still frame from the same video sequence. A box in the mosaic indicates the individual frame.



A B

Figure 1-1. Video Mosaic Image and single extracted still frame.

## 1.2 **Background**

As observed in the Gulf War, space-based surveillance assets and manned platforms alone could not satisfy the war fighter's desire for continuous, on-demand, situational awareness information [7]. This trend continued during operations in Bosnia. While space-based or manned surveillance systems produced excellent imagery, the process of scheduling collection platforms or gaining access to the information proved too time consuming. At times, the imagery collection assets were not available nor in the proper position to obtain the required imagery of a target. The frustration over the inability to obtain the required information spawned an effort to develop a theater controlled, imagery reconnaissance platform capable of long endurance, enabling coverage of a typical operating theater. The

result was a system called the medium altitude endurance unmanned aerial vehicle (MAE-UAV) which is now referred to as "Predator" [23].

The information provided by the Predator system consists of real-time full-motion video (FMV), synthetic aperture radar (SAR) to allow imaging through cloud cover, and infrared motion imagery for collection of FMV at night. Predator operators provide modest "triage level" exploitation of the FMV at its ground control station (GCS). The exploitation of the FMV performed in the GCS produces a limited number of still images exported directly to upper echelon intelligence headquarters [23].

During the 1997 deployment to Bosnia in the former Republic of Yugoslavia, the Predator system was augmented with intelligence personnel and equipment housed in a facility known as the rapid exploitation and dissemination (RED) cell. The RED cell produces frame-grabbed still images and intelligence reports that are more in-depth and detailed than the near real-time still images and moderate triage-level analysis provided by the predator operators in the GCS [23].

### 1.3 **Problem**

The United States military has responded to the lessons learned in previous armed conflicts, such as the Gulf War and operations in Bosnia, with an increased focus on UAV reconnaissance platforms and the products they provide. The resulting explosion in the volume of UAV FMV has created a huge videotape library. The imagery data contained on those tapes is valuable to analysts who are tasked with providing intelligence reports for tactical war fighters. The enormous volume of video data creates a huge burden for analysts who are often searching for only a small number of scenes. In addition, UAVs continue to

collect streams of FMV, which if stored on tape will only add to the problem. Therefore, it is important to provide a method for storing and retrieving the information contained in FMV without needing to view the entire tape or stream.

The use of UAV imagery provides the war fighter with increased situational awareness in the theater of operation. Access to the analog tapes is limited to those who have physical custody of them. Making the UAV imagery available to the war fighter in the trenches will increase their awareness and ultimately their effectiveness. As discussed earlier, intelligence analysts store both frame-grabbed still images and reports for subsequent retrieval on a imagery server. The still frames extracted by GCS operators and intelligence analysts represent a small fraction of the total information contained in the UAV FMV. A more complete representation of the UAV FMV is needed that can be easily accessed and browsed. Accordingly, the focus of this research is to identify the steps required to transform UAV FMV into a form that can be accessed by users at all levels, especially those attempting access over low-bandwidth tactical communications systems. A demonstration incorporating web and video mosaic technology is utilized to determine the feasibility of the proposed approach.

### 1.4 **Research Objective**

The objective of this research is to develop a process employing the video mosaic building technology that converts video data into a still image format without losing content of the video stream. A commercially developed application is employed to build mosaic images from sequences of still image files extracted from a video stream. The proposed

mosaic building process provides detailed steps for converting a video stream to a series of still images that can be used as indexes to the larger video segments.

### 1.5 **Scope**

Limitations of the current level of the individual technologies used in this research, such as video segmentation, video mosaic building, and meta data extraction prevent complete automated tool integration. Thus, process development is performed manually for this research effort. This section describes the scope of the individual components of the process proposed in this research.

- *Imagery Collection*. This research focuses on the full motion video (FMV) imagery provided by the Predator UAV system. Video imagery used in this research is actual reconnaissance data provided by Air Force Research Laboratory Information Directorate (AFRL/IF), from the Expeditionary Force Exercise (EFX) 98 held at Eglin AFB, Florida, 14-16 Sep 98.
- *Imagery Type*. This research will focus on Daylight visual FMV because it has the most immediate face-value.

### 1.6 **Approach and Presentation**

A process development and implementation approach is being utilized in this research. A survey of previous work is provided in Chapter Two to provide the user knowledge of key concepts and formats necessary to understand subsequent chapters. Chapter Three presents a detailed explanation of the process developed for this research effort. Chapter Four presents and discusses the results of a manual implementation of the

process on a sample video stream. The conclusions and recommendations are presented in Chapter Five.

# 2 SURVEY AND ANALYSIS OF PREVIOUS WORK

## 2.1 Introduction

As described in Chapter One, the steadily growing archive of surveillance videotapes makes locating video segments based on a user's specific criteria increasingly difficult. Consequently, this research effort will determine if video segmentation and mosaic building techniques can be favorably applied to unmanned aerial vehicle (UAV) full motion video (FMV) for alternate data representation.

In an effort to help the reader better understand the methodology used in this research, a basic understanding of the Predator system, imagery formats, video manipulation, and video storage techniques are presented. Section 2.2 provides a basic description of the Predator UAV system, which is the imagery collection platform of choice for this research. Next, the image and video formats are described in Section 2.3. Section 2.4 and 2.5 provide surveys of video stream segmentation and video mosaic technologies, respectively. Finally, Section 2.6 describes video storage and retrieval.

## 2.2 Predator System

The Predator system was designed to fill a void in reconnaissance imagery products experienced by local theater commanders during Desert Storm [7]. This section presents a high level description of the predator system to give the reader a general understanding of the flow of the reconnaissance data. The data provided by this system is of extreme value to the war fighter.

### 2.2.1 *Overview*

The Predator aircraft is an unmanned aerial vehicle (UAV) whose reconnaissance payload collects imagery data that is transmitted to the GCS via a data link. A pictorial representation of the Predator system is shown in Figure 2-1. The current Predator system is made up of two to four Predator aircraft, a ground control station (GCS), Trojan Spirit II SATCOM communication systems, and ground support equipment (GSE) [23]. Each

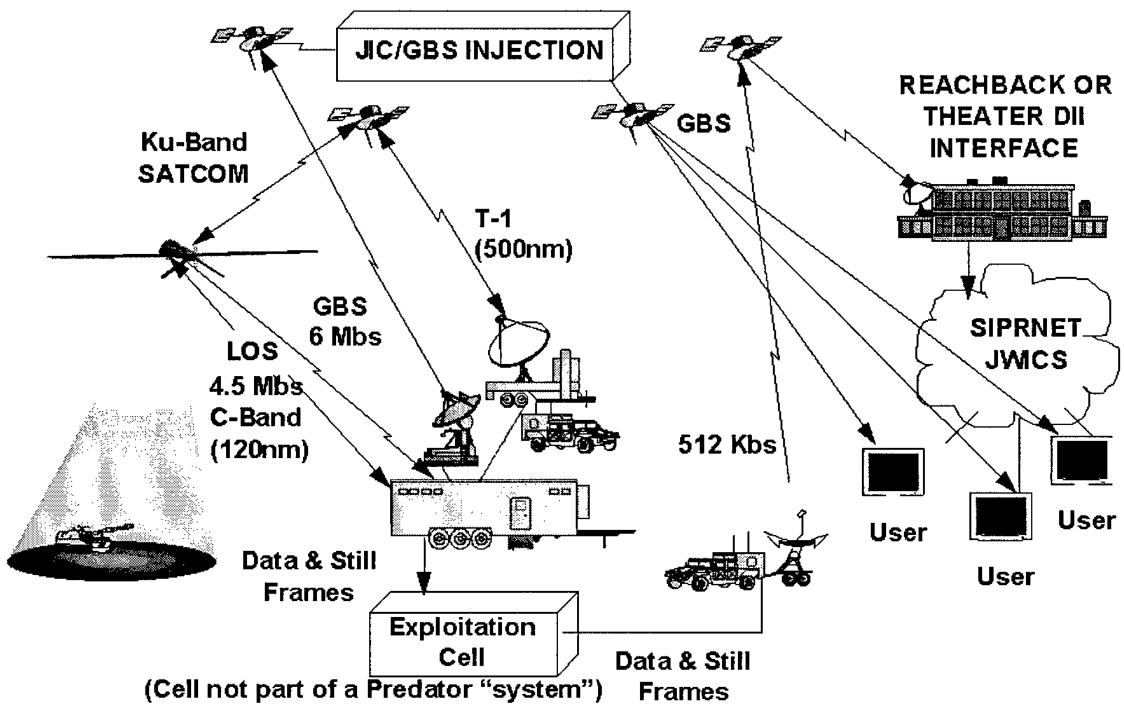



Figure 2-1. Predator System Overview [8]

subsystem performs a crucial part of the imagery data collection and delivery process.

### 2.2.2 *Predator Aerial Vehicle*

The aircraft used in the Predator system is a mid-wing monoplane with an inverted V-tail. A Predator aerial vehicle in flight is shown in Figure 2-2. It can be controlled remotely using a data link to provide control commands, or it can be operated autonomously by

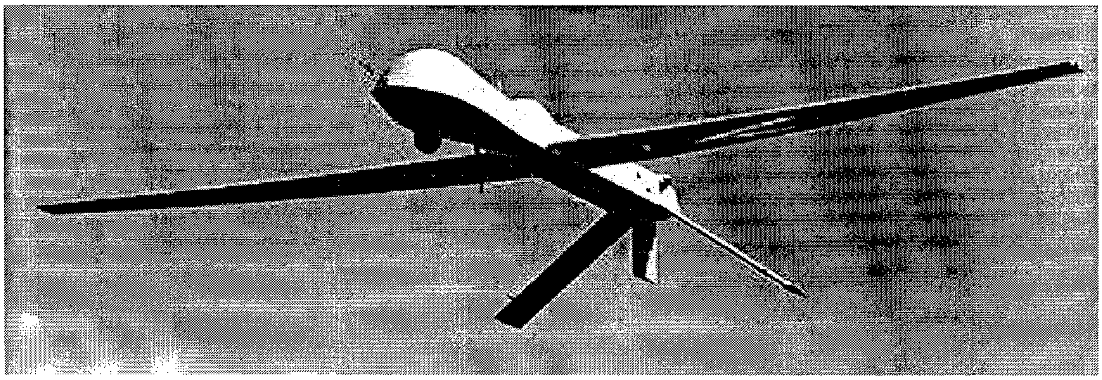


Figure 2-2. Predator Unmanned Aerial Vehicle in flight.

executing a preprogrammed mission. The aircraft is capable of carrying Daylight TV (DLTV), Electro-Optical/Infrared (EO/IR) and Synthetic Aperture Radar (SAR) payloads, all of which provide its surveillance capability. The DLTV camera collects full motion video (FMV) in the visible light spectrum, while the EO/IR payload produces FMV data in EO/IR electro-magnetic spectrum [9].

### 2.2.3 *Summary of Predator System*

The Predator system provides the capability for medium altitude reconnaissance using an unmanned aerial vehicle. The payload of the UAV can benefit tactical forces by providing intelligence information about a target objective or operations area. Once the imagery data has been collected by the Predator system it can then be transmitted to a field location or command center for analysis.

## 2.3 Image and Video Formats

There are a number of standard image and video formats that will be described in this section to enable the reader to understand their use in subsequent sections.

### 2.3.1 *Joint Photographic Experts Group (JPEG)*

The name "JPEG" has been used to refer to a compression technique based on the standards developed by the Joint Photographic Experts Group. The group combines expertise in television engineering, computer science, and many other disciplines to focus on human vision and computer graphics. The resulting JPEG technique uses subsampling and quantization to selectively identify and remove information to which the human eye is less sensitive. Subsampling reduces the resolution of the color information by one-half. Quantization is achieved through rounding the discrete cosine transform (DCT), which is a mathematical representation of the pixel difference within a 8 x 8 pixel block. The result is considered "lossy" because the decompressed image is not identical, pixel for pixel, to the original image; however, the differences are not visually noticeable. The fact that the JPEG technique discards some information every time the image is compressed makes it a poor candidate for detailed imager analysis (e.g. sub-pixel analysis), however, the high compression offered by the JPEG format makes it ideal for image archival [14].

### 2.3.2 *Portable Pixel Map (PPM)*

The PPM format is a simple graphics format for storing color images. The format consists of a header followed by a list of the pixels contained within the image. This format does not incorporate compression or special encoding. The pixels in the image are represented either in binary representation or as ASCII decimal numbers. The individual color components (red, green, and blue) stored in binary each take one byte. This binary format can only handle up to 256 (0 to 255) color levels per component. The color components stored in ASCII decimal are not limited to 256 color levels. The pixels are written row by row from top to bottom and are each written as red, green, and blue values,

respectively [14]. The simplicity of the PPM file format is countered by its bulky nature due to a lack of compression.

### 2.3.3 *Thumbnail*

A thumbnail is a reduced size/resolution representation an image [14]. The size of the images depends greatly on the dimensions of the original image. In addition to reducing the size of the image, the resolution is reduced to 72 pixels/inch. For example, the thumbnail for a 16 Kilobyte (KB) image (400 x 231 at 120 pixels/inch) is approximately 3KB (96 x 54 at 72 pixels/inch). The pixel width of the original image, the standard width of the thumbnail, and the change in resolution, all determine the reduction factor. The thumbnail size and resolution are configurable by the user.

### 2.3.4 *Moving Picture Experts Group (MPEG)*

The Motion Picture Experts Group was organized to develop standards for high quality video compression. The name "MPEG" is also used to refer to the video compression techniques based on the standards developed by the motion pictures experts group. The standards define compression methods for both video and audio. The original video format based on this standard is MPEG-1. MPEG-1 supports television quality video (i.e., 30 frames per second). The bit rate of the MPEG-1 data stream is 200 kilobytes per second, and the quality is comparable to VHS videotape. The MPEG-2 video format supports high quality video and requires high-speed digital connectivity from 1.5 to 2.5 megabytes per second. MPEG-2 is closely related to high definition television (HDTV). The MPEG formats combine a number of compression techniques similar to those described in the JPEG format along with techniques for encoding differences between successive frames. MPEG stores four different kinds of frames: I-frames, B-frames P-frames, and D-frames. I-frames,

also known as *independent* frames or key frames, do not require any additional information to decode. I-frames are compressed using the same general technique as JPEG compression. P-frames, or *predictive* frames, actually contain the difference from the previous I-frame or P-frame. MPEG uses a method employing motion prediction to store P-frames as offsets for 8 X 8-pixel squares. B-frames or *bi-directional predictive* frames use differences from both previous and future frames. D-frames are separate, low-resolution versions of frames, which are similar to thumbnails, that are intended to simplify browsing. D-frames are rarely used. The I-frames contain all the information needed for decoding; however, P-frames and B-frames depend on other frames before they can be decompressed. A sequence of MPEG frames is shown in Figure 2-3. The arrows point to frames, which must be decoded prior to the frame originating the arrow. For example, frame 2 requires information from frame 3 before it can be decompressed, frame 3 needs frame 0, and frame 6 requires that frame 3 be decompressed first. These types of dependencies can require the compressed frames to be stored in the file out of order to ensure that frames needed by other frames are decompressed first. One possible order for the frames shown in Figure 2-3 is 0, 3, 1, 2, 4, 6, 5 [13].

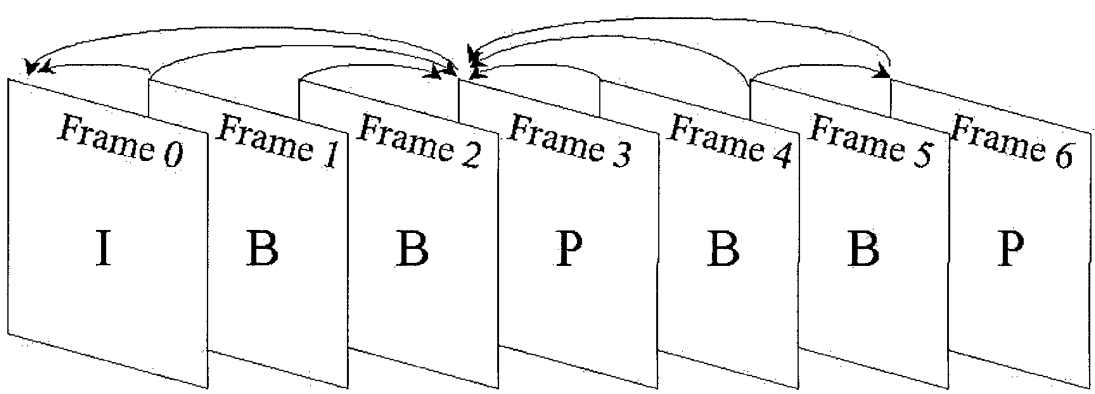



Figure 2-3. A Sequence of MPEG Frames[14].

## 2.4 Video Stream Segmentation

A video stream collected from a reconnaissance platform can be several hours in length. A single second of MPEG video can reduce thirty 68.8 (352 x 240) kilobyte frames, totaling over 2 Megabytes, to only 200 kilobytes (compression level is based on selected compression criteria). Even compressed, a video stream of considerable length in the MPEG-2 video format will result in a very large file. Hence, a two-hour video with 344 x 200 pixel frames could encode to an MPEG-2 file as large as 288 (1440 sec x 200 kilobytes/sec) Megabytes. A video file of this size needs to be divided into manageable segments that can be processed into mosaics (section 2.5) and/or stored in a video database (section 2.6), which are covered in the next two sections.

Effective event-based video segmentation can be accomplished by locating scene breaks within the video stream. A scene break occurs when the difference between two subsequent frames exceeds a predefined threshold of a given characteristic. In recent years, there has been considerable attention placed on scene change detection. Two of the methods used for scene change detection are histogram-based (Sethi & Patel)[20] and motion-based (Bhandarkar & Khombhadia) [1]. Histogram-based scene change detection employs intensity histograms, and more recently color histograms, to calculate the difference measure from two adjacent frames. Motion-based parsing of video uses block-based motion compensation and the discrete cosine transform (DCT)-coded prediction error signal. Both of these techniques are having success in detecting scene changes, however, only recently has scene change detection technology been applied to surveillance video [19]. The utility of this

technology stems from the need for organizing video, or in the case of this research, mosaic building.

## 2.5 **Building a Video Mosaic Image**.

The video mosaic image combines the information from a contiguous series of video frames, while eliminating redundant information. This essentially converts from a video format to a still image format, while retaining information from the entire video sequence. The result is a smaller file requiring fewer resources to process or transmit. In addition, video mosaics provide a more complete representation of a video sequence as compared to single frame-grabbed images. Several mosaic-building applications are available, such as Frame Stitcher™ by Litton© Corporation, VideoBrush Panorama™ by VideoBrush© Corporation, and Visual Stitcher™ by PanaVue©. All applications were designed to operate under the Microsoft© Windows™ environment. Following, is a brief description of all three applications followed by an explanation of video characteristics that cause them problems.

### 2.5.1 ***Frame Stitcher***

The Frame Stitcher GUI Application [12], developed through the Air Force Research Laboratory Information Research Directorate, performs batch processing of sequential frames in PPM format. The Frame Stitcher application aligns the overlapping portions of frames using the Heath algorithm. The Heath algorithm uses varying degrees of resolution to fine-tune the alignment process. Figure 2-4 depicts how overlapping frames can be aligned using a process that begins with a very coarse resolution representation of two overlapping frames.

The process continues using progressively finer resolutions to achieve a "best match", see Figure 2-4. Once all files to be included are processed, the mosaic image is complete. The Frame Stitcher application offers two options for exporting the mosaic image: 1) as an

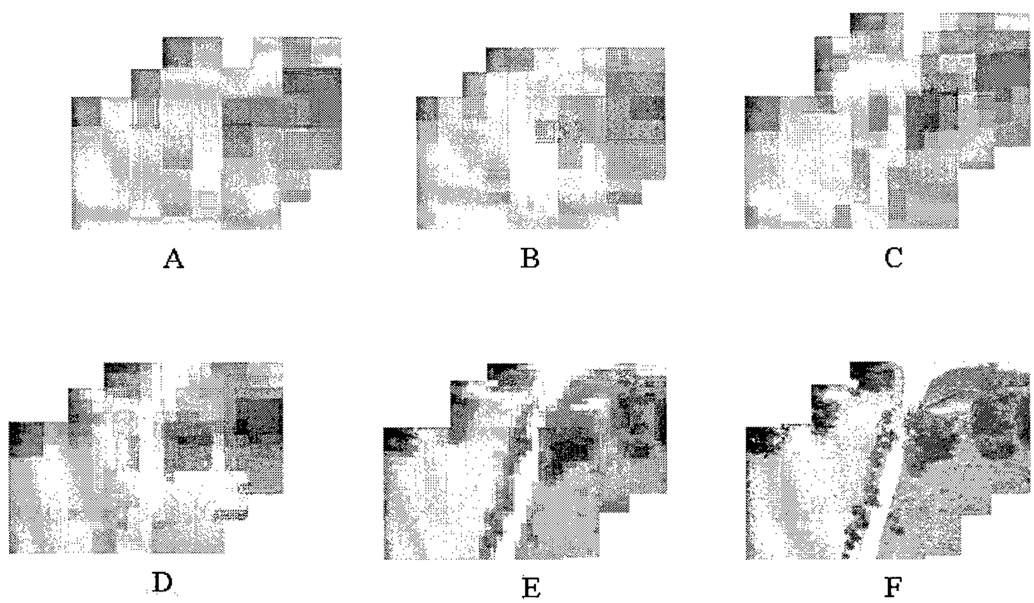



Figure 2-4. Overlapping Image Alignment Using Varying Resolutions.

image in PPM format, or 2) a data file in mosaic information file (MIF) format. The PPM file is a composite image of all frames processed through the Frame Stitcher application. The MIF file is a data file containing the name of each frame file used to create the mosaic and the pixel offset values for each frame, see Figure 2-5. The Frame Stitcher GUI was developed for the Microsoft Windows environment.

```
frame_0=vfile=sc19_40.ppm drawn=true x=0 y=0
frame_1=vfile=sc19_41.ppm drawn=true x=0 y=0
frame_2=vfile=sc19_42.ppm drawn=true x=-1 y=-2
frame_3=vfile=sc19_43.ppm drawn=true x=-1 y=-3
frame_4=vfile=sc19_44.ppm drawn=true x=-1 y=-4
frame_5=vfile=sc19_45.ppm drawn=true x=-2 y=-5
frame_6=vfile=sc19_46.ppm drawn=true x=-2 y=-6
frame_7=vfile=sc19_47.ppm drawn=true x=-2 y=-7
frame_8=vfile=sc19_48.ppm drawn=true x=-2 y=-8
frame_9=vfile=sc19_49.ppm drawn=true x=-2 y=-10
frame_10=vfile=sc19_50.ppm drawn=true x=-2 y=-12
```

Figure 2-5. Example Frame Stitcher Mosaic Information File (MIF).

### 2.5.2 *VideoBrush Panorama*

The VideoBrush Panorama™ application [22] creates a still image video mosaic file from either, 1) Video for Windows AVI format files, or 2) live video from standard Video for Windows compatible devices. Image files can be output in either Windows Bitmap (BMP) or JPEG (JPG). The VideoBrush Panorama application was developed for the Microsoft Windows environment.

### 2.5.3 *Visual Stitcher*

The Visual Stitcher application [18] stitches a row or column of photos into a single panorama. Visual Stitcher accepts photos in the following formats: BMP, JPG, TIF, TGA, PCX, PSD, PIC, PCD, and FPX. Visual Stitcher can save a project file that identifies the files stitched or the actual panoramic image. Images can be output in the following formats: BMP, JPG, TIF, TGA, PCX, PSD, PCD, FPX, PCT, PNG and MOV. Visual Stitcher stitches the input photos based on overlap placement determined solely by the user, or on overlap placement determined using application assistance.

### 2.5.4 *Video Characteristics that Hamper Mosaic Building*

There are many characteristics of UAV reconnaissance video that make some segments non-candidates for mosaic building. The problem traits consist of zooms, tilting, or scene changes/bad frames. A brief description of each is provided.

#### 2.5.4.1 *Zoom*

A zoom can occur through sensor actions or platform actions. The FMV sensor can increase/decrease magnification to provide an enlarged/wide angle view of the sensor aim point. The zoom effect can also occur by movement of the sensor, and reconnaissance platform, in respect to the sensor aim point, i.e., a change in the range from target. For example, if the range between the Predator aircraft and the sensor aim point were to steadily decrease or steadily increase, while the magnification on the sensor remained constant, the effect is the same as a zoom in or out, respectively.

#### 2.5.4.2 *Tilting*

A tilting of the scene occurs when the orientation of a scene rotates about a point in the sensor field of view as shown in Figure 2-6. A tilting occurs when the sensor platform circles a target or the sensor changes orientation with the reconnaissance platform itself.

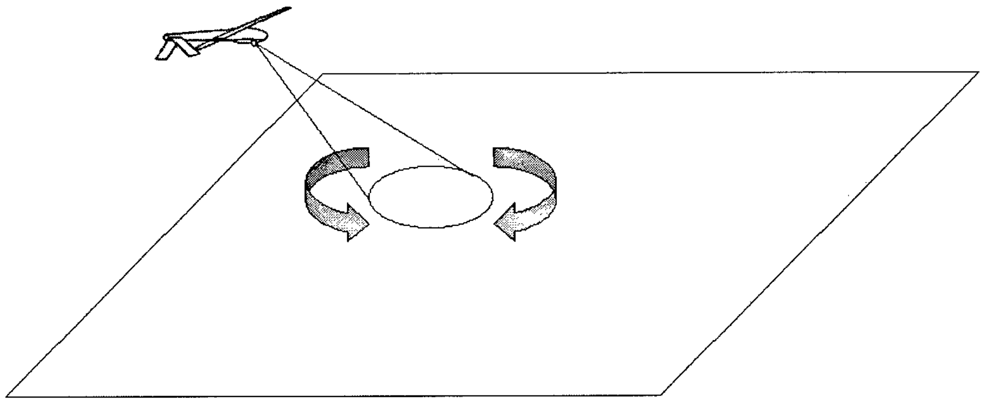
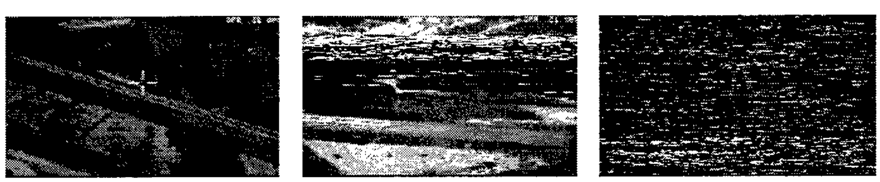


Figure 2-6. A Depiction of Scene Tilting Effect .

2.5.4.3 *Scene Changes/Bad Frames*

A scene change occurs when there is a significant change between sequential frames in a video stream. A bad frame occurs when either glare (poor exposure) or noise is introduced into the video image. In Figure 2-7A, the image is poorly exposed by the sensor due to either glare or a lack of ambient daylight. Figure 2-7B and C are examples of moderate to extreme noise, respectively. Scene changes and bad frames have both been successfully detected using the segmentation techniques described earlier in this chapter.



A B C

Figure 2-7. Three levels of video degradation.

2.5.4.4 _Summary_

Video sequences having any of these traits are categorized as non-candidates for building a video mosaic based on the current level of mosaic technology. All applications discovered during this research and discussed earlier in this chapter are designed to create mosaic images from sequences of images that do not contain any of the problem traits just discussed. Video mosaic images built from sequences exhibiting any of these traits have unpredictable results.

## 2.6 Video/Image Storage and Retrieval

Multimedia data such as the reconnaissance video and associated meta data (e.g. longitude, latitude, range to target, date/time) collected by the Predator UAV can be stored in multimedia databases. A video database management system provides efficient access to archived digital video. The Air Force Image Products Library (IPL) uses an object-oriented approach for the database and standard web technology for both data access and presentation [16]. This section explains each of these areas to help the reader understand how the proposed mosaic building process will fit into existing technology.

### 2.6.1 ***Video Database Management System (VDBMS)***

The goal of a VDBMS, like its traditional text-based counterpart, is to make retrieving data stored in the database both convenient and efficient. The VDBMS, however, must tackle the complicated task of retrieving video [10]. The method proposed by Yeo and Yeung [24] handles video as a multimedia object defined using a video hierarchy. The hierarchy uses a clip-scene-shot structure. The terms used in the VDBMS hierarchy correspond to the video stream, segment, and selected frames in this research. The goal of

the video hierarchy is to overcome the sequential and time-consuming process of viewing video. Figure 2-8 shows the concept of video hierarchy as it corresponds to this research. The video hierarchy shown in Figure 2-8 depicts how the mosaic, or its associated thumbnail, can be used as an index to the particular video segment.

### 2.6.2 *Visual Information Retrieval (Browsing)*

Recent advances in information retrieval systems provide the capability to retrieve

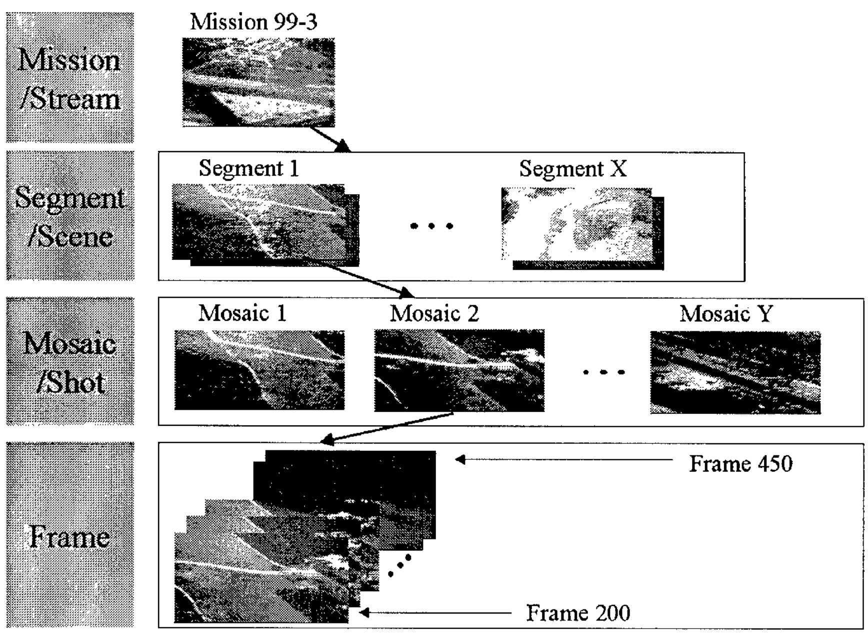



Figure 2-8. A Hierarchy of Digital Video.

digital images and videos using examples and/or visual sketches. The visual properties of the imagery, such as colors, textures, shapes, motions, and spatiotemporal compositions, are used in combination with text and other related information. The items that are returned from searches using these methods are not exact matches but rather a "best match" [2, 3].

### 2.6.3 ***Object Oriented Database***

An object-oriented approach is well suited for multimedia applications because of the ability to define multimedia data types [4]. In addition, the Object Management Group (OMG) has developed a commercially accepted Common Object Request Broker Architecture (CORBA) that provides the basic administration client, task management, toolkits and services for the U.S. Air Force's Virtual Distributed Library (VDL) [16].

### 2.6.4 ***World Wide Web (WWW) Technology***

The power and utility of the WWW makes it an excellent choice to provide access to JPEG-formatted image data anywhere in the internet-connected world. The standard applications, such as web browsers, servers, and search engines, that have been developed to leverage the strengths of the web, provide a rich set of tools to enhance data retrieval. The major components required to provide access to data are explained below.

#### 2.6.4.1 *Web Server*

A web server contains the host software required to answer requests for data. The host processes the requests for files contained on the local system and replies with the requested information [21].

#### 2.6.4.2 *Web Client*

A web client, also referred to as a browser, provides the user an interface to view hypertext markup language (HTML) pages. The browser allows the user to request files from other machines (web servers) located anywhere on the WWW [21].

#### 2.6.4.3 *HyperText Markup Language (HTML) Page*

HTML enriches text documents with a variety of markup, making it possible to transfer virtually any type of data. HTML specifies the general appearance of a text

document and can contain links to other documents. A file referred to by a link could be local or anywhere on the internet. The linked file could be an image, a video file, or another web page [21].

2.6.4.4 *Active Server Pages (ASP)*

Active server pages are programmable web pages mixing HTML, ODBC commands, and scripting code. The ASP technology developed by Microsoft© provides the capability to manipulate the contents of a database and dynamically generate web pages to present the results of a database query [21].

## 2.7 **Summary**

The Predator system collects imagery information valuable to operational theater commanders. The thousands of hours per year of data acquired by the Predator system become part of a large image repository. Organizing the video imagery is very challenging due to its streaming nature. Advances in the area of video segmentation and mosaic building, combined with a hierarchical design for organizing the components of the video by-products, will allow more efficient access to the video information. Once the video is processed into a hierarchical structure, it can be searched and accessed via the WWW.

# 3 VIDEO MOSAIC BUILDING PROCESS

## 3.1 **Introduction**

Reconnaissance platforms such as the Predator UAV produce a constant stream of MPEG-2 video data. The reconnaissance video stream is currently archived on analog tape. The accessibility of UAV imagery data is limited due to its analog format and large size. Archived video data is currently contained on thousands of tapes. Once an analyst locates a tape, they must perform painstaking frame-by-frame analysis of UAV video to find a particular target or scene of interest. Consequently, the access to the imagery data is limited to those who have access to the imagery tapes and the equipment required to view them. The fast-paced tempo of military operations and geographically separated operating locations require that the imagery data be converted to a more portable and accessible form, compatible with reduced bandwidth capabilities associated with the tactical communications environment.

The most logical approach to provide this capability is to place the UAV video into a VDBMS. As described in Chapter Two, the video stream must be segmented, organized, and indexed to allow piecewise access to FMV scenes or scene content. Serving this information over the Internet makes it available to the widest possible audience.

The remainder of this chapter focuses on the detailed steps required to capture, organize, and store UAV video data. A series of web pages are then used for browse, search, and retrieval of the video data, associated still images, and meta data. The resulting process

converts video data into a still image format to be served over the internet, see Figure 3-1, for a graphic depiction of the entire process.

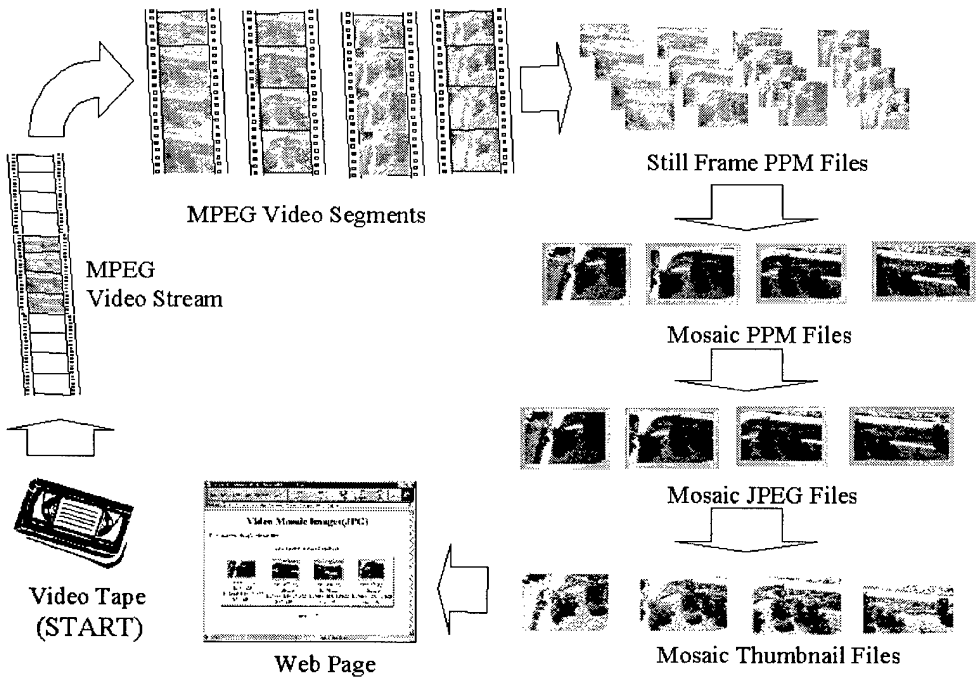



Figure 3-1. Overall Conversion Process (Video/Mosaic/Thumbnail).

## 3.2 **Process Description**

This section presents a step-by-step breakdown of the entire process. Each step contains input data, some type of processing, and output data. The output of a step flows logically to the input of the following step.

### 3.2.1 ***Data Collection***

The video stream is either processed directly or captured from tape and stored in MPEG-2 format on a computer accessible medium such as magnetic disk.

### 3.2.2 *Identify Video Sequences to Build Mosaics*

To qualify as a candidate sequence for building a mosaic, the video sequence must be a contiguous sequence of frames free of problem frames or camera actions as described in Chapter Two. Figure 3-2 shows an MPEG-2 video stream that is divided into segments. Segment boundaries can be event-based as described in Chapter Two or temporal-based.

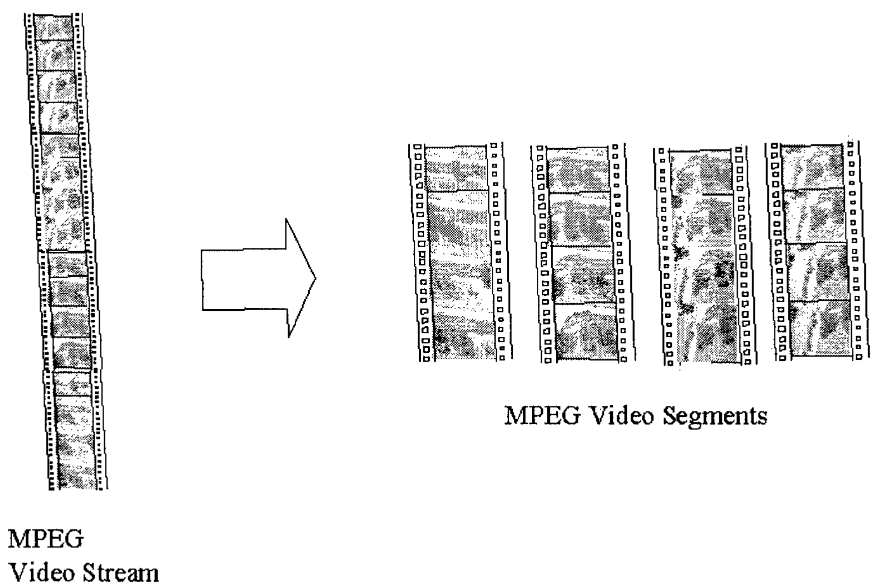



Figure 3-2. Segmentation of MPEG-2 Video.

Temporal-based sequences result when a video sequence exceeds a predetermined time limit. Establishing a temporal limit maintains the goal of files, both still image and video, which are efficiently downloadable over low-bandwidth communication lines. The resulting segments are identified as either video mosaic candidate or non-candidate sequences as described in Chapter Two.

### 3.2.3 *Identify Frame to Represent Non-Candidate Video Sequences*

Still frames are needed to represent the video sequences identified as non-candidate sequences for building a mosaic. As described in Chapter Two, some video sequences cannot be built into a mosaic. To represent these video segments, a frame is selected at regular intervals throughout the problem segment. The rate of frame selection is set to ensure there are no gaps in coverage for the segment. Still frames selected are converted to a compressed format such as JPEG for archival.

### 3.2.4 *Decode MPEG-2 Video Stream*

The MPEG-2 video stream is decoded and individual video frames are stored in a still image format such as PPM, see Figure 3-3. The frames are grouped according their location in the stream and are identified with either a mosaic candidate or non-candidate sequence.

During decoding, meta data tagged onto the video stream, and meta data associated with the video sequence, or individual frames can be extracted and associated with the proper

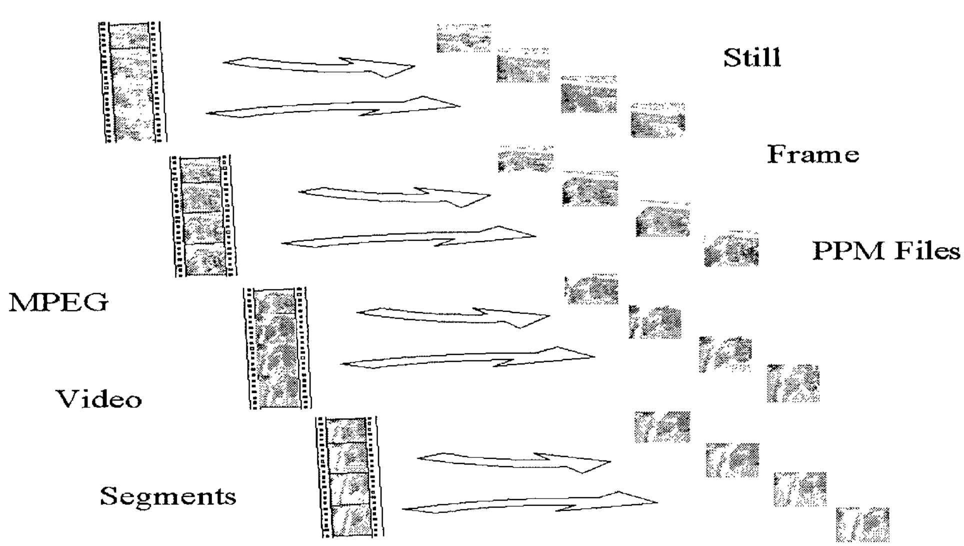



Figure 3-3. MPEG Segments Decoded into Still Image Files.

frames. It is possible to encode meta data directly onto the audio channel of the MPEG-2 stream. This method creates an efficient way of "piggy-backing" meta data on the video stream without degrading the video image. Handling meta data is discussed in more detail later in this chapter.

### 3.2.5 *Build Mosaics*

The frame image files from the mosaic candidate sequences are processed through a mosaic-building application in sequential order. The resulting video mosaic is a single image and is stored in a still image format such as PPM, see Figure 3-4. Once the mosaic images are created, they can be converted to an archival format such as JPEG and made accessible from the VDBMS server.

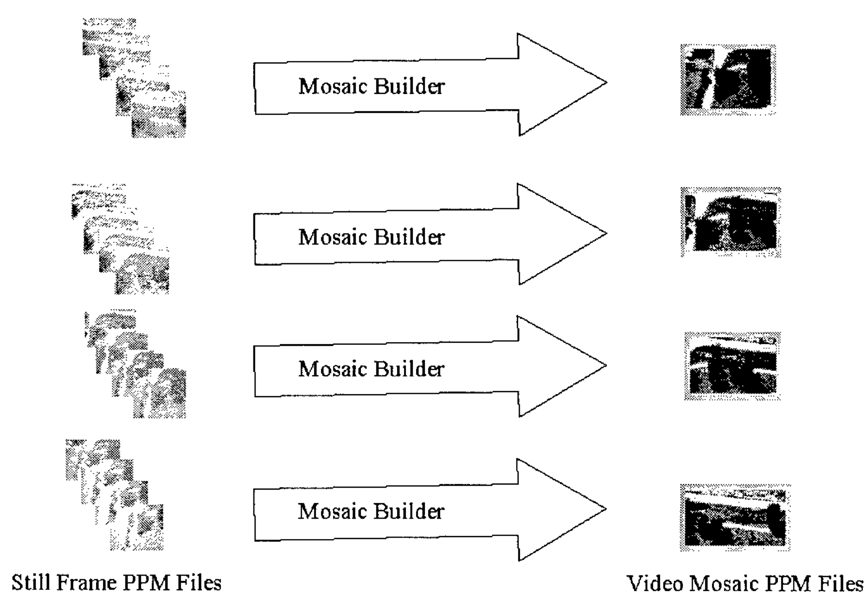


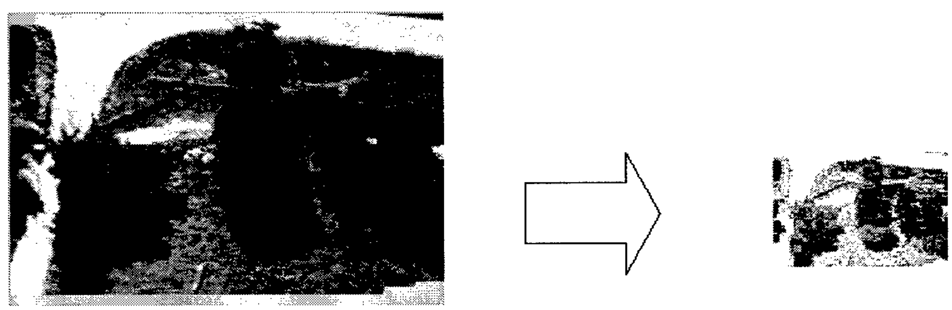

Figure 3-4. Creation of Video Mosaic a using Sequence of Still Frames.

### 3.2.6 *Extract Thumbnails for Mosaics and Selected Still Images*

Thumbnail images are produced for each mosaic image and the selected still frames. An industry accepted thumbnail creation application converts an image to a lower resolution, e.g. from 344 x 200 at 120 pixels/inch to 96 x 52 at 72 pixels/inch, as described in Chapter Two (see Figure 3-5). The loss in image fidelity is acceptable when compared to the greatly reduced transmit time. The thumbnail images are embedded in web pages as a low-resolution preview of the mosaic and selected frame images.

### 3.2.7 *Collect Meta Data for Video Stream and By-Products*

Meta data is collected for the MPEG-2 video segments and still images. The meta



A B

Figure 3-5. JPEG image Converted to Thumbnail Image.

data can be collected either manually or automatically. Manual entries will comprise information only attainable through direct observation by an analyst, such as recognition of a small or specific detail in the photo not detectable through pattern matching techniques. Meta data can also be collected manually from the "burned in" or *reticulated* data on the video. Automated collection can be accomplished in a number of ways, and it will be

utilized for the majority of the meta data collection. Meta data can be extracted from the UAV FMV as an asynchronous digital feed (e.g., MPEG-2 private data channel), from the closed caption segment of the FMV, from the audio channel, or through optical character recognition (OCR) from reticulated data on video [11]. Possible meta data associated with the UAV FMV consists of longitude, latitude, and altitude of the UAV, latitude and longitude of the sensor aim point, sensor aim point range from UAV, time into video, Julian time, width of sensor field of view, etc. In addition, meta data can be gathered from alternate sources. For example, data contained in a database may have information about a building or other structure located at a particular set of coordinates. If the sensor aim point in the UAV FMV contains the same coordinates, the information pertaining to the structure could become meta data for the UAV video and associated mosaic image. Meta data is extremely important, because it forms the foundation by which the majority of the searches over the archived data will be based. Meta data are indexible items.

### 3.2.8 *Place Archival Files on Data Server*

The MPEG-2 segments, mosaic images, thumbnails, and associated meta data are stored in a format consistent with the hierarchical structure of the VDBMS described in Chapter Two. The meta data is entered with the object it is associated with and proper links are inserted in the database to provide pointers to the actual physical files (MPEGs & JPEGs). Maintaining the scene-level organization when storing the mosaic images provides direct access to individual video scenes.

### 3.2.9 *Create Static Web Pages*

The thumbnail images, meta data, and links to MPEG and JPEG files are displayed via static web pages as described in Chapter Two. The images are placed in the static web page in sequential order, as shown in Figure 3-6.

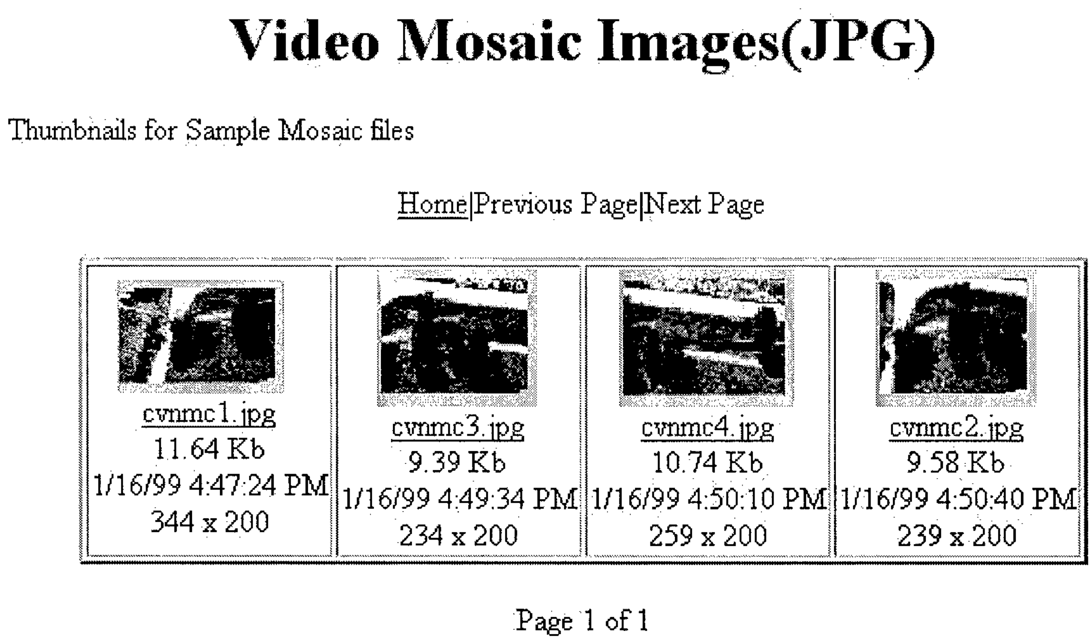



Figure 3-6. Static Web Page Displaying Mosaic images

### 3.2.10 *Deploy Web Search Page*

A search page is positioned on the web server with appropriate links to the VDBMS. The search page collects search criteria from the user and builds a query to execute against the database. The link to the database will allow the search page to pass query commands and receive/format, query results. The meta data discussed earlier provides a rich choice of query opportunities that can be used to greatly narrow the search, returning a small set of

specific results. A smaller set of results reduces transmit times which is in alignment with the goal of this research.

### 3.2.11 ***Serve Web Pages on Demand***

The web pages are positioned on a web server. The web server provides connectivity to the Internet and executes script codes embedded within the web pages to facilitate database searches and dynamic web page creation.

## 3.3 **Summary**

The focus of this research is to combine video and WWW related technologies into a process, which will convert UAV video, stored on analog videotape into a form that facilitates online storage and rapid search/retrieval. This process consists of eleven steps starting with capturing video from analog tape through serving a still image representation of the video, with browse and search capabilities, over the Internet.

# 4 PROCESS IMPLEMENTATION AND DEMONSTRATION

## 4.1 Overview

This chapter describes an implementation of the proposed video mosaic-building process, as described in Chapter Three, on a representative video stream. The researcher will describe specific details pertinent to the implementation. In addition, the researcher will use the description of the Predator system as described in Chapter Two of this paper to determine a logical location for the proposed mosaic building process, and associated image manipulation processes.

## 4.2 Process Implementation and Demonstration Development

This section uses the step-by-step process description found in Chapter Three to process imagery data provided by AFRL/IF.

### 4.2.1 *Data Collection*

Data collection began with the receipt of an 8-millimeter videocassette provided by AFRL Information Directorate, which contained UAV reconnaissance video footage. The analog signal was captured from the tape using a SONY© HI-8 player and SNAZZI™ [6] video capture hardware and software at the 88th Communications Group Multimedia Center, Wright-Patterson AFB, OH. Once captured, the video segment was converted to MPEG-2 format. The MPEG-2 format was selected because it is the format the Predator system uses to transmit its reconnaissance video. This process was repeated several times to acquire a number of video streams that were representative of actual UAV mission video footage.

### 4.2.2 *Identify Video Sequences to Build Mosaics*

The MPEG-2 video files must be analyzed to Identify mosaic candidate sequences as described in Chapter Three. A standard MPEG player, which labeled each frame number and the time into the stream, was used to view the MPEG-2 file. This was a tedious process and required frame-by-frame analysis to identify the traits described in Chapter Two that cause problems for mosaic building applications. A 35.3-second MPEG file containing 1057 frames was selected as representative of UAV video and contained both mosaic candidate and non-candidate sequences, including zooms, tilts, and reticulations. A total of 27 mosaic candidate sequences were identified based on criteria stated in Chapter Two.

### 4.2.3 *Identify Frames to Represent Non-Candidate Video Sequences*

The Non-candidate sequences contained in the video stream were analyzed. As described in Chapter Three, frames should be captured from non-candidate segments to fill in between the mosaic candidate segments to ensure there are no gaps in the still image representation of the video stream. The degree of overlap was analyzed and it was determined that every tenth frame, one-third second, a frame should be selected. A total of 40 frames were identified to represent 21 segments.

### 4.2.4 *Decode MPEG-2 Video Stream*

As described in Chapters Two and Three, the MPEG-2 stream must be decoded into individual frames before it can be processed by the mosaic building software. An MPEG-2 decoder called *mdcdeco* developed by the Computer Science Department at Wayne State University [15] was used to decode the video stream. The *mdcdeco* application was selected

because it did an excellent job decoding the captured MPEG files, it is freeware, and PC-based. The frames were saved in the PPM format, which is the format needed by the mosaic builder (see 4.2.5).

The original file size of the MPEG-2 video file was 17,993 Kbytes and decoded to 262 Mbytes (1057 frames x 248Kbyte frame size). The MPEG-2 file was re-encoded using the cropped PPM files resulting in a 9,921 Kbyte file. All reduction percentages will be based on the MPEG created from the cropped images.

### 4.2.5 ***Build Mosaics***

The Frame Stitcher application was provided by the sponsor of this research and neither VideoBrush Parnorama nor Visual Stitcher tested better using actual UAV imagery. Brief production descriptions are provided in Sections 2.5.1 through 2.5.3.

As described in Chapter Two, the Frame Stitcher application accepts a file that contains a list of sequential frames to be built into a mosaic image. The Frame Stitcher application reads the file and one-by-one processes the PPMs.

During the course of this research it was necessary to crop the reticulated data prior to building the mosaic. This was necessary because this data obscured the mosaic image and resulted in a poor product. An ImageMagick [5] *crop* utility was used, in conjunction with a DOS batch routine, to semi-automatically crop the images. Eight pixels on the left and 40 pixels on the top of each frame were cropped resulting in a 20 percent pixel count reduction.

Once the reticulated data was cropped, Frame Stitcher could be used to build the mosaic images. Each image was saved in PPM format. Each mosaic was then converted to the JPEG format for archival. The ImageMagick *convert* utility was used in conjunction with a DOS batch file to semi-automatically convert the PPM files to the JPEG format. Frames selected to fill the gaps between the mosaic candidate segments are also converted to the JPEG format for archival. The 67 mosaic and selected still images combine for a total size of 764 Kbytes.

#### 4.2.6 *Extract Thumbnails for Mosaics and Selected Still Images*

The Thumbs Plus™ application created by Cerious Software Inc. was inexpensive, and quickly and easily created thumbnail images from selected files or for an entire directory. The default width of 96 pixels was used and deemed sufficient through observation of thumbnails extracted from the mosaic and selected still images. The combined size of the 67 thumbnails representing the mosaic and still images is 199 Kbytes, 26 percent of the original JPEGs sizes.

#### 4.2.7 *Collect Meta Data for Video Stream and By-Products*

As described in Chapter 3, meta data associated with the video stream is important as a source of information and for use in querying a database. In this research, a set of data was created to simulate meta data for the example video stream. All meta data values were created to be semantically accurate, however they are for demonstration purposes only. Meta data is described in Section 3.2.7.

### 4.2.8 *Place Archival Files on Data Server*

Once the image files are in archival format, they are ready to be stored on the data server. As described in Chapter Three, the mosaic identifier is entered into the database with associated meta data and pointers to physical file locations. For this research, the data is stored in a Microsoft© Access 7.0 database. This database along with Windows 98 open database connectivity (ODBC) drivers provided sufficient functionality for data storage and retreival.

### 4.2.9 *Create Static Web Pages*

The static web pages are created using the thumbnails for the mosaic and selected images, meta data for each image, and links to the physical files. The ThumbsPlus application was used to create the basic structure of the static web page and Microsoft's FrontPage Express™ was used to update the attributes beneath each image, see Figure 4-1.

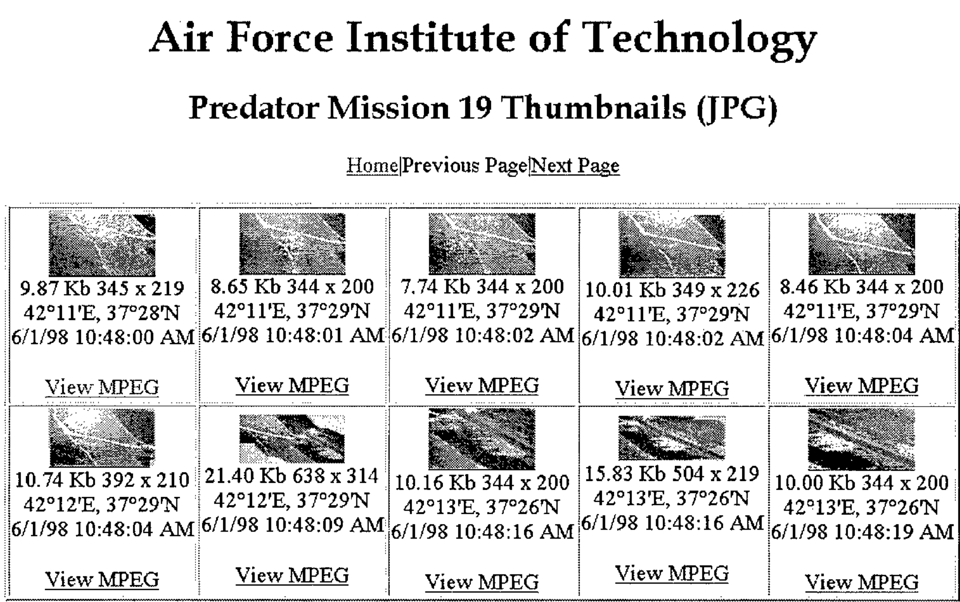
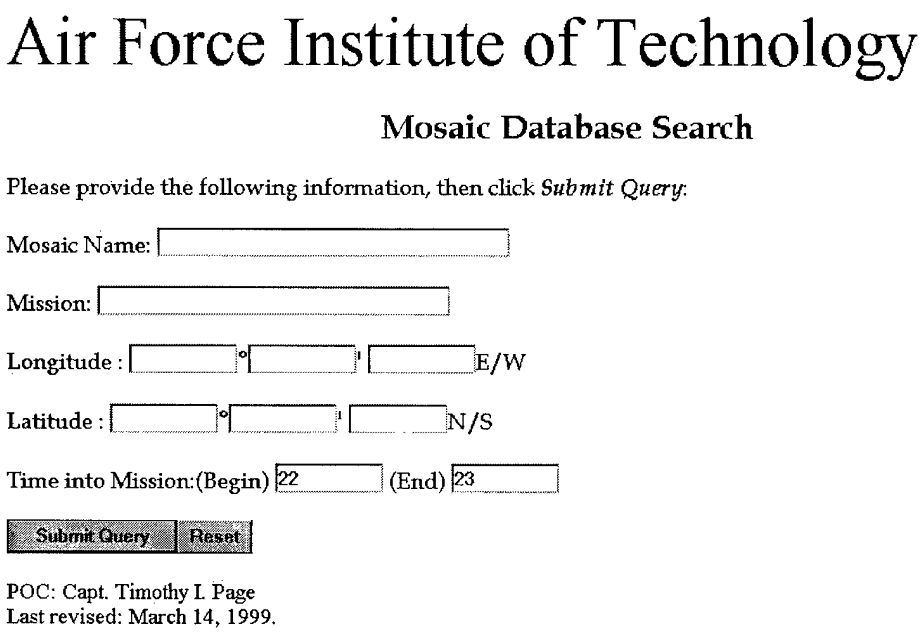



Figure 4-1. Static Web Page for Browsing Mosaics.

FrontPage was also used to add links pointing to the physical video segment files. This allows direct retrieval of the video file from the static page.

#### 4.2.10 *Deploy Web Search Page*

The web search page collects information the user would like to search on, see Figure 4-2. As described in Chapter 3, the search page uses this information to build a query and



Air Force Institute of Technology

**Mosaic Database Search**

Please provide the following information, then click *Submit Query*:

Mosaic Name:

Mission:

Longitude : ° ' E/W

Latitude : ° ' N/S

Time into Mission:(Begin) 22 (End) 23

Submit Query Reset

POC: Capt. Timothy L Page
Last revised: March 14, 1999.

Figure 4-2. Query Criteria Collection Form.

then processes the query against the database. This process results in a page being dynamically created using the information retrieved from the database, see Figure 4-3. The search criteria collection page and the form used to format query results were both created using Microsoft's FrontPage Express.

# Air Force Institute of Technology

## Imagery Database

(Search Results)

| | | | Longtude | | | Latitude | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Mosaic Name | Size | Mission | min | sec | dir | min | sec | dir | Time into Mission | Mosaic Hyperlink | Video Hyperlink |
| mc19_13.jpg | 344X200 | Mission 19 | 42 | 14 | E | 37 | 25 | N | 22 |  | Click to View Video Segment |
| mc19_14.jpg | 344X200 | Mission 19 | 42 | 14 | E | 37 | 25 | N | 22 |  | Click to View Video Segment |
| mc19_15.jpg | 344X200 | Mission 19 | 42 | 14 | E | 37 | 25 | N | 23 |  | Click to View Video Segment |

The information mentioned herein is fictitious and is in no way intended to represent any real events unless otherwise noted.

Figure 4-3. Search Results Page.

### 4.2.11 *Serve Web Pages on Demand*

Microsoft's Personal Web Server™ was used to serve the web pages. This application provided a cost-effective solution for serving the web pages created for this thesis effort.

### 4.2.12 *Summary*

Actual UAV FMV was processed using the steps described in Chapter Three. The process converted the 9,921 Kbyte video stream into 67 still images totalling 764 Kbytes. Thumbnail representations of the still images totaling 199 Kbytes, were served as low-resolution previews for the actual still images using web pages. In addition, a search capability was provided allowing the user to retrieve selected images/data from the database/image archive. All data and products associated with this research are available as outline in Appendix A.

## 4.3 Video Mosaic Process Location in the Predator System

The proposed process uses UAV reconnaissance data as input. It is important to retrieve the video data at a point that leverages benefits of the proposed process. This section will evaluate the Predator system data path as described in Chapter Two, to determine a logical location for the proposed video mosaic and web page building process.

### 4.3.1 *Ground Control Station*

The reconnaissance video travels from the UAV to the ground control station (GCS) where technicians currently perform triage level data analysis. The GCS would be a good location for equipment used in the mosaic building process due to the availability of non-reticulated data. If a remote image product library (IPL) server could be installed in the GCS, this would make the GCS a logical choice to locate the proposed process. If, however, a remote IPL server could not be installed in the GCS, this placement would add a large amount of data transmission requirements to the link between the GCS and the Joint Analysis Center (JAC) where the 5D (see below) database is stored. In addition to the video data, still images, meta data, and web pages would need to be transmitted.

### 4.3.2 *Archival Location*

Currently, all still frames extracted from the full motion video at the GCS are stored in the national imagery archive housed in the Demand Driven Direct Digital Dissemination (5D) database at Molesworth, England [16]. The most likely storage location for the video mosaic and video segment files is the 5D database at Molesworth. Co-locating the video mosaic process with the final storage site eliminates the need to transmit all of the processed files over communication links. Co-locating the mosaic building process and storage

hardware provides a logical solution because it adds minimal communication requirements. Processing the FMV reconnaissance data through the segmentation, mosaic building, thumbnail, and web page building processes produces a large amount of data for archival. For example, 2 hours of MPEG-2 video with the same characteristics as the stream processed above would equate to over 2 Gbytes of data just for the MPEG itself. Processing the MPEG would result in an additional 9792 segments, yielding over 13,000 images totaling over 155 Mbytes. In addition, the thumbnail images add another 40 Mbytes resulting in a grand total of over 2.2 Gbytes of digitized imagery. The user can access the information in the 5D database through the IPL interface located at Molesworth where the FMV reconnaissance data arrives in near-real time. Unfortunately the FMV currently arriving at Molesworth is reticulated data, consequently the additional lossy effort of image cropping would be incurred.

### 4.3.3 *Process Location Conclusion*

All of the previously mentioned factors make co-locating the video mosaic-building process at the archival site a logical choice. A possible archival site is the national imagery database in Molesworth, England. However, the current non-availability of non-reticulated at Molesworth degrades this solution. Another possible choice is the GCS. If a remote IPL database and server could be installed in the GCS, the added availability of non-reticulated data would make the GCS the logical choice. In addition, the GCS is normally located in close proximity to the frontline users needing the information, thereby cutting down on possible propagation delay.

## 5 CONCLUSIONS AND RECOMMENDATIONS

### 5.1 Introduction

This chapter uses the products of the implementation of the proposed video mosaic building process, as described in Chapter 4, as a basis for the following conclusions and recommendations.

### 5.2 Conclusions

On the basis of the results obtained during the implementation of the proposed video mosaic building process, the following conclusions are drawn:

- It is possible to implement the proposed process on unmanned aerial vehicle (UAV) full motion video (FMV). The implementation of the proposed process using UAV FMV produced a sequence of images that provided a favorable representation the original reconnaissance video stream.
- The video mosaic images created by the process proposed in this research are more efficiently retrievable and viewable than the video streams themselves due to the smaller file size and panoramic presentation of the information. The proposed process produced a representation of the video stream with a favorable reduction in size. The implementation of the proposed process on the UAV FMV produced a set of mosaic and selected still images with a combined size much smaller than the original compressed video. In addition, the use of thumbnails to represent the mosaic and selected still images further reduces the size of the represented information for preview purposes.
- The reduced file size resulting from the implementation of the proposed process equates to a larger user set having connectivity to access the mosaic representations of the video

segments. The larger user set results from the reduced transmission requirements needed to effectively access and download the information.

- The future usefulness of the proposed process will rely heavily on the availability of a non-reticulated video stream. If non-reticulated data is not available, the process will need to crop the reticulated data. This step could eliminate valuable information. The current operation of the Predator system does not forward non-reticulation video past the GCS. There are two possible solutions:
    1) Forward the non-reticulated video stream to the location of the mosaic building process. According to this research sponsor, since the development of the subject process, new UAV footage production includes non-reticulated images. All telemetry and meta data are transmitted synchronously and separately from the images themselves. Implementation of this format to the Predator UAV system will ensure that non-reticulated data is available to the mosaic building process. However, all legacy footage includes the reticulated data.
    2) Co-locate the mosaic building process, a remote IPL database server and the GCS.
- The development of automatic scene change detection will play an important role in the success of automating the creation of video mosaic images. The increased ability to organize and index at the scene level make scene change detection an important part of managing video data. In addition, the detection of scene changes is crucial to identifying video segments that cannot be processed through the video mosaic application.

- Based on the amount of additional data generated, a logical placement of the proposed mosaic building process in the Predator system is at the archival site.

## 5.3 Recommendations

Based on observations during the development and implementation of the proposed process, the following recommendations are proposed for further research:

- Further research and development could be accomplished to automate the process as the technology of mosaic building and scene change detection matures. The proposed process uses publicly available and inexpensive software making future research into automation attractive to future academic researchers.
- The provision of non-reticulated video could be accomplished through the implementation of non-destructive methods for attaching meta data to a UAV video stream. Current research into non-destructive methods for attaching meta data to a video stream as indicated by AFRL/IFE, should be implemented into the Predator UAV System.

## 5.4 Summary

The process developed in Chapter Three showed favorable outcomes based on products and observations of the implementation. The near-term benefit is a method for converting a UAV video stream into a hierarchical structure, which uses video mosaic images as indexes to segments within the video stream. This process can be placed in the UAV system at the imagery archival location to avoid increasing current transmission levels. Also, the organization created by the video hierarchy leads to reduced download requirements when browsing or searching the database. To realize these benefits, it is recommended that

future research focus on the automation of the proposed process and implementation of a non-destructive method for attaching meta data to MPEG video.

# Appendix A – DATA AND SOFTWARE AVAILABILITY

The data and software used in this research is available by contacting the AFIT School of Engineering Database Systems Research Point of Contact (POC). Currently, the Database Research POC is:

Major Michael L. Talbert
Air Force Institute of Technology
WPAFB, OH 45433-7765

Email: michael.talbert@afit.af.mil
Phone: DSN 785-6565 ext. 4280 COMM (937) 255-6565

## BIBLIOGRAPHY


[1] - Bhandarkar, S. & Khombhadia, A., "Motion-based Parsing of Compressed Video". IEEE, June 1998:80-87, 1998.

[2] - Bouet, Marinette & Djeraba, Chabane, "Visual Content Based Retrieval in an Image Database with Relevant Feedback", IEEE Computer 0-8186-8676-6/98, 1998.

[3] - Chang, Shih-Fu, Smith, John R., Beigi, Mandis & Benitez, Ana. "Visual Information Retrieval from Large Distributed Online Repositories", Communications of the ACM, December 1997/Vol. 40. No. 12.

[4] - Christodoulakis, Stavros & Koveos, Leonidas, "Multimedia Information Systems: Issues and Approaches". *Modern Database Systems* by Won Kim, ACM Press, ISBN 0-201-59098-0, 1995:318-337.

[5] - Cristy, John, "ImageMagick Image Manipulation Software", http://www.wizards.dupont.com/cristy/ImageMagick.html, 1996.

[6] - DAZZLE™ Multimedia, "SNAZZI Video Capture Software", http://www.dazzlemm.com.sg/html/snazzi.html, 1996.

[7] - Defense Airborne Reconnaissance Office (DARO), "Unmanned Aerial Vehicle (UAV) 1995 Annual Report", http://www.fas.org/irp/agency/daro/uav95/other.html, 1995.

[8] - "Endurance UAV - Concept of Operations", http://www.fas.org/irp/doddir/usaf/conops_uav/, 1996.

[9] - General Atomics - Aeronautical Systems Inc., "Predator Aerial Vehicle Operation Manual", ASI-00023, *Prepared for the Naval Air Systems Command, Contract NO. N00019-96-C-0121*, CDRL sequence No. D016, 1997.

[10] - Grosky, William I. "Managing Multimedia Information in Database Systems", *Communications of the ACM*, December 1997, vol. 40. no. 12.

[11] - Hansen, Steve, "Automated Spatio-temporal Archiving and Retrieval of Unmanned Air Vehicle Full Motion Video": Draft Document, The MITRE Corporation, 1998.

[12] - Heath, Tom, "Frame Stitcher Graphical User Interface (GUI) Application User's Manual". Air Force Research Laboratory, Information Directorate, Rome NY, 1997.

[13] - Kay, David C. & Levine, John R. "Graphics File Formats", Published by Windcrest™ of McGraw-Hill, 1995.

[14] - Kientzle, Tim "Internet File Formats", Coriolis Group Books, 1995.

[15] - Li, Dongge & Sethi, Ishwar K., "MPEG-2 Decoding and Encoding Software", http://vision.cs.wayne.edu, Vision and Neural Networks Laboratory, Computer Science Department, Wayne State University, Detroit, MI 48202, 1998.

[16] - National Imagery and Mapping Agency (NIMA), "Image Product Library", Briefing Slides, *http://www.if.afrl.af.mil*, Air Force Research Laboratory, Information Directorate, Rome NY, 1999.

[17] - Patel, Nilesh & Sethi, Ishwar K., "Compressed Video Processing For Cut Detection", Vision and Neural Networks Laboratory, Computer Science Department, Wayne State University, Detroit, MI 48202, 1997.

[18] - PanaVue, "Visual Stitcher Image Stitching Software", http://www.panavue.com/, 1999.

[19] - Pyburn, Bradly, "Analysis of the Applicability of Video Segmentation to Unmanned Aerial Vehicle Surveillance Video", Air Force Institute of Technology, GCS-ENG-99M-17, 1999.

[20] - Sethi, Ishwar K. & Patel, Nilesh, "A Statistical Approach to Scene Change Detection", Vision and Neural Networks Laboratory, Computer Science Department, Wayne State University, Detroit, MI 48202, 1995.

[21] - Stout, Rick, "The World Wide Web, Complete Reference", Osborne McGraw-Hill, ISBN 0-07-882142-8, 1996.

[22] - VideoBrush Incorporated, "VideoBrush Panorama creation software", http://www.videobrush.com/ , 1999.

[23] - Wiedemann, P. , "On The Use of The Predator (MAE-UAV) System in Bosnia", as *Presented at the Unmanned Vehicle Conference on 13 Jun, 1997 in Paris, France.*

[24] - Yeo, Boon-Lock & Yeung, Minerva M. "Retrieving and Visualizing Video", *Communications of the ACM*, December 1997/Vol. 40. No. 12.

## BIOGRAPHICAL SKETCH

Captain Timothy I. Page is a prior enlisted Air Force communications officer from Millerton, PA. He attended Williamson Jr. and Sr. High School. Captain Page earned an associate degree in Computer Science from Williamsport Area Community College before enlisting in the Air Force as a System Repair Technician for the Air Force Technical Applications Center (AFTAC) in 1983. He spent ten months at Lowry AFB for technical training and was assign to Technical Operations Directorate, Maintenance division, at McClellan AFB. Captain Page volunteered as part of the elite honor guard for McClellan AFB and was selected as first ever honor guard of the quarter. While at McClellan, Captain Page was awarded a full scholarship to the University of Wilkes-Barre where, in 1990, he earned a BS in Computer Science.

Captain Page returned to active duty in 1990 attending the Basic Communications-Computer Officer Training program. He was then assigned to the Air Force Operational Test and Evaluation Center (AFOTEC) as a test director. He developed a plan for testing large software systems that became the AFOTEC standard. Captain Page was then assigned to NorthEast Air Defense Sector at Griffiss AFB, NY. There he acted as Chief Quality Assurance Branch and later Chief of Operations.

In 1994, Captain Page was assigned to the 88th Communications Group, Wright-Patterson AFB, OH. He was matrixed to the F-16 System Program Office where he acted as Chief of Management Information Systems (MIS) Branch. As Chief MIS, he was responsible for a $4 million budget to develop a wide area network between the F-16 SPO at WPAFB, OH and its logistics counterpart at Hill AFB, UT. The F-16 SPO boasted the top network on WPAFB and was responsible for breaking new ground and setting the standard. He was selected Company Grade Officer of the Quarter for October-December 1995 and Company Grade Officer of the Year for 1995.

Captain Page is currently assigned to the Air Force Institute of Technology (AFIT), WPAFB OH, where he is pursuing a Master degree in Computer Systems. Captain Page's military awards include the Commendation Medal (third oak leaf cluster), Air Force Good Conduct Medal, National Defense Service Medal, and the Air Force Organizational Excellence Award (second oak leaf Cluster). After graduation in March 1999, Captain Page has been selected to instruct at the Communication-Computer Officers Training School at Keesler AFB, MS. Captain Page is married to the former Carol Honabach of Wilkes-Barre, PA, and has three children: Gregory, Kevin, and Christyn.